



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/513,058	02/24/2000	Anita Wai-Ling Huang	AM999074	5519
23334	7590	07/30/2003		
FLEIT, KAIN, GIBBONS, GUTMAN & BONGINI, P.L. ONE BOCA COMMERCE CENTER 551 NORTHWEST 77TH STREET, SUITE 111 BOCA RATON, FL 33487			EXAMINER	
			BASEHOAR, ADAM L	
		ART UNIT	PAPER NUMBER	
		2178		6

DATE MAILED: 07/30/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

P24

Office Action Summary	Application No.	Applicant(s)
	09/513,058	HUANG ET AL.
Examiner	Art Unit	
Adam L Basehoar	2178	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) Responsive to communication(s) filed on 24 February 2000.

2a) This action is **FINAL**. 2b) This action is non-final.

3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) Claim(s) 1-26 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) Claim(s) _____ is/are allowed.

6) Claim(s) 1-26 is/are rejected.

7) Claim(s) _____ is/are objected to.

8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) The specification is objected to by the Examiner.

10) The drawing(s) filed on _____ is/are: a) accepted or b) objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11) The proposed drawing correction filed on _____ is: a) approved b) disapproved by the Examiner.

If approved, corrected drawings are required in reply to this Office action.

12) The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

13) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a) All b) Some * c) None of:

1. Certified copies of the priority documents have been received.

2. Certified copies of the priority documents have been received in Application No. _____.

3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

14) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).

a) The translation of the foreign language provisional application has been received.

15) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

1) Notice of References Cited (PTO-892)

2) Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) Information Disclosure Statement(s) (PTO-1449) Paper No(s) 2.

4) Interview Summary (PTO-413) Paper No(s) _____.

5) Notice of Informal Patent Application (PTO-152)

6) Other: _____

DETAILED ACTION

1. This action is responsive to communications: Information Disclosure Statement filed on 02/24/00 of the application filed on 02/24/00.
2. Claims 1-26 are pending in the case. Claims 1, 10, 12, 17, and 25 are independent claims.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 1-9 and 12-24 are rejected under 35 U.S.C. 103(a) as being unpatentable over Brown et al (US: 5,913,208, 06/15/99).

-As per independent claim 1, Brown et al teach a method for identifying duplicate documents, where having *received said documents* generating a hit-list (Fig. 3b) of intrinsic and non intrinsic attributes (title, author, creation date, length, size, location, abstract, etc (columns 1 and 6, lines 64-67 and 51-54)) wherein the table structure (equivalent to metadata summary subtrees) of the attributes of each document pair are compared (i.e. if one document is missing an attribute structure then the documents are considered distinct) and each document is shown to be

distinct from the other if their respective attributes aren't equal (Fig. 4: 430 & 450). Brown et al do not disclose that the summary hit-list contain metadata. It was well known in the art that document attributes such as title, size, location, etc are considered metadata of a document. Brown et al also do not disclose that the summary hit-list includes a sub-tree structure. It would have been obvious to one of ordinary skill in the art, to have used Brown et al method of comparing hit-lists of attributes and used the structural relations of those hit-list attributes as the first determinate of distinct documents, because while both methods provide the same benefit of identifying distinct documents, by checking the structure (layout) of those attributes first, it can more quickly be established with the minimum amount of comparisons that the compared documents are distinct of one another (i.e. both documents have structures A & B, but since one has A before B and the other B before A they are considered distinct.)

-As per dependent claim 2, Brown et al further disclose wherein the table structure contains intrinsic and non intrinsic *attributes* (title, author, creation date, length, size, location, abstract, etc (columns 1 and 6, lines 64-67 and 51-54)) to be compared.

-As per dependent claim 3, Brown et al further disclose that one of the attributes used for determination could be an abstract or a title (text-content)(column 6, lines 51-54).

-As per dependent claim 4, Brown et al further disclose and that one of the attributes used for determination could be an abstract or a title (text-content)(column 6, lines 51-54).

-As per dependent claim 5, Brown et al further disclose noting the current pair of hit-list documents as duplicates if the text content are *equal* (i.e. the structure and non-text attributes have been compared but alone can not determine duplicates) (Fig. 4: 460).

-As per dependent claim 6, Brown et al further disclose one embodiment of *deleting duplicates* from the hit list group (column 8, lines 20-22: Fig. 8).

-As per dependent claims 7-9, Brown et al further disclose that in the hit-list (metadata table) when an entry is identified as having one or more duplicates.....is cross referenced in the duplicate identifier field of the other duplicate (column 8, lines 11-16), wherein for two metadata entries the *entry number and duplicate identifier field constitute a 2x2 matrix* (Fig. 3b). Brown et al also teach that the process of identifying duplicates as distinct is achieved by comparing the metadata structures, attributes, and text content of the attributes as seen in claims 1-3. Brown et al do not teach storing a zero binary value in the first row and second column position of the matrix to designate that the first and second documents are distinct. It would have been obvious to one of ordinary skill in the art, to have used a binary indicator (0 or 1) in Brown et al duplicate identifier position concerning only two documents, because while Brown et al is set up for a plurality of documents which could lead to many multiple dependencies, if a restriction of only comparing two documents was applied to Brown et al, there would be no need for more than one indicator (0 or 1) to clearly state the relationship between the first document and second document.

-As per independent claim 12, Brown et al teach a system wherein a *search engine process produces a hit-list* (metadata parser) (Abstract), where it has already been stated above that the hit-list contains metadata (i.e. structure, attributes, and some text content), wherein a hit-list *formatter processor selects one or more of.... duplicate instances of one another* (summary consolidator)(Abstract) with the ability of the formatter processor to delete duplicates from the hit list group (repository) (column 8, lines 20-22: Fig. 8). Brown et al also teach wherein the generated summary hit-lists are stored (in a repository) (Fig. 1 and Fig. 4: 410)

-As per dependent claim 13, Brown et al further teach that the formatter processor (summary consolidator)(Fig. 4: 420 & 440) is configured to compare the metadata attributes values of the hit-lists (metadata summaries) and compare the text content (i.e. abstract or title) included in the metadata attribute hit-lists. As shown in claim 1, Brown et al does not show that its formatter is configured to compare the structures of the metadata hit lists. It would have been obvious to one of ordinary skill in the art, to have used Brown et al method of comparing hit-lists of attributes and used the structural relations of those hit-list attributes as the first determinate of distinct documents, because while both methods provide the same benefit of identifying distinct documents, by checking the structure (layout) of those attributes first, it can more quickly be established with the minimum amount of comparisons that the compared documents are distinct of one another (i.e. both documents have structures A & B, but since one has A before B and the other B before A they are considered distinct.)

-As per dependent claims 14-16, Brown et al further teach wherein the system is configured to compare the attribute values included in the hit-lists, the text content included in the metadata hit-lists, and the metadata portion of the hit-list.

-As per independent claim 17 and dependent claims 18 and 19, Brown et al teach the method of classifying electronically posted documents of claim 1-3 respectively. Brown et al does not teach that the method is a program product to be executed by a computer system stored in a computer readable medium. It was well known in the art to convert a method for use by a computer system to a program product so that that method can be executed on said system and any other systems that can execute the program. It was also well known in the art that a computer system for executing said program product would have a recordable media (memory).

-As per dependent claim 20, Brown et al further teach identifying documents as duplicates if text content of the metadata hit-lists (i.e. abstract or title) are equivalent (i.e. the structure and non-text attributes have been compared but alone can not determine duplicates) (Fig. 4: 460).

-As per dependent claim 21, Brown et al further teach removing the duplicate metadata hit-list from the group (column 8, lines 20-22; Fig. 8).

-As per dependent claims 22-24, Brown et al further teach the hit-list (metadata table) when an entry is identified as having one or more duplicates.....is cross referenced in the duplicate

identifier field of the other duplicate (column 8, lines 11-16), wherein for two metadata entries the *entry number and duplicate identifier field constitute a 2x2 matrix* (Fig. 3b). Brown et al also teach that the process of identifying duplicates as distinct is achieved by comparing the metadata structures, attributes, and text content of the attributes as seen in claims 1-3. Brown et al do not teach storing a zero binary value in the first row and second column position of the matrix to designate that the first and second documents are distinct. It would have been obvious to one of ordinary skill in the art, to have used a binary indicator (0 or 1) in Brown et al duplicate identifier position concerning only two documents, because while Brown et al is set up for a plurality of documents which could lead to many multiple dependencies, if a restriction of only comparing two documents was applied to Brown et al, there would be no need for more than one indicator (0 or 1) to clearly state the relationship between the first document and second document.

Claims 10-11 and 25-26 are rejected under 35 U.S.C. 103(a) as being unpatentable over Brown et al (US: 5,913,208, 06/15/99) in view of Microsoft Press Computer Dictionary, Microsoft Press, 1997, pp. 309.

-As per independent claim 10, Brown et al teach the method of claim 1 as stated above, and further disclose receiving not just two, but *two or more entries* (column 3, lines 54-55) of documents and generating metadata hit-list summaries for each received document. Brown et al do not teach that when receiving a plurality of documents to group the metadata hit-list summaries according to the mime-type designation of each document. The Microsoft Press Computer Dictionary teaches us that MIME-types describe the contents of a document which can be used to interpret the content of a file over the internet (pp. 309). It would have been obvious

to one of ordinary skill in the art, to have used Brown et al method for identifying duplicate documents from search results without comparing document content and grouping the hit-list summaries by MIME-type, because documents of different MIME-types wouldn't have to be compared as their intrinsic attributes would be wholly different, i.e. a regular text/plain MIME-type couldn't be a duplicate of a text/html MIME-type, and thus this would increase efficiency by significantly reducing the number of hit-list summary comparisons.

-As per dependent claim 11. Brown et al and the Microsoft Press Computer Dictionary do not teach grouping more subsets of hit-list summaries by MIME-type. As it was taught above in claim 10, it would have been obvious to have grouped as many MIME-type subsets as were required by the plurality of documents for the purpose of increased efficiency by significantly reducing the number of hit-list summary comparisons across MIME-types.

-As per independent claim 25, Brown et al teach the method for classifying electronically posted documents of claim 10 as stated above. Brown et al do not teach that the method is a program product to be executed by a computer system stored in a computer readable medium. It was well known in the art to convert a method for use by a computer system to a program product so that that method can be executed on said system and any other systems that can execute the program. It was also well known in the art that a computer system for executing said program product would have a recordable media (memory).

-As per dependent claim 26. Brown et al does not teach grouping more subsets of hit-list summaries by MIME-type. As it was taught above in claim 25, it would have been obvious to have grouped as many MIME-type subsets as were required by the plurality of documents for the purpose of increased efficiency by significantly reducing the number of hit-list summary comparisons across MIME-types.

Conclusion

5. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

US 6547829	04/15/03	Meyerzon et al.
US 6269362	07/31/01	Broder et al.
US 6487553	11/26/02	Emens et al.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Adam L Basehoar whose telephone number is (703) 305-7212. The examiner can normally be reached on M-F: 7:30am - 4:00pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather Herndon can be reached on (703) 308-5186. The fax phone numbers for the organization where this application or proceeding is assigned are 703-746-7239 for regular communications and 703-746-7238 for After Final communications.

Art Unit: 2178

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.

ALB

July 28, 2003



Heather R. Herndon
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100